

# 1

## Introduction

There is a revolution brewing in ecology. Granted, it is a gentle and slow revolution, but there is growing dissatisfaction with the statistical methods that have been most commonly taught and used in ecology (Hilborn and Mangel, 1997; Wade, 2000; Clark, 2005).<sup>1</sup> One aspect of this revolution is the increasing interest in Bayesian statistics (Fig. 1.1). This book aims to foster the revolution by making Bayesian statistics more accessible to every ecologist.

Ecology is the scientific study of the distribution and abundance of biological organisms, and how their interactions with each other and the environment influence their distribution and abundance (Begon *et al.*, 2005). The discipline depends on the measurement of variables and analysis of relationships between them. Because of the size and complexity of ecological systems, ecological data are almost invariably subject to error. Ecologists use statistical methods to distinguish true responses from error. Statistical methods make the interpretation of data transparent and repeatable, so they play an extremely important role in ecology.

The Bayesian approach is one of a number of ways in which ecologists use data to make inferences about nature. The different approaches are underpinned by fundamentally different philosophies and logic. The appropriateness of different statistical approaches has been fiercely debated in numerous disciplines but ecologists are only now becoming aware of this controversy. This occurs at least in part because the majority of statistical books read by ecologists propound conventional

<sup>1</sup> The conventional statistical methods are known as frequentist statistics and include null hypothesis significance testing (NHST) and construction of confidence intervals. NHST attracts the most criticism. See Chapter 2 for more details of these methods.

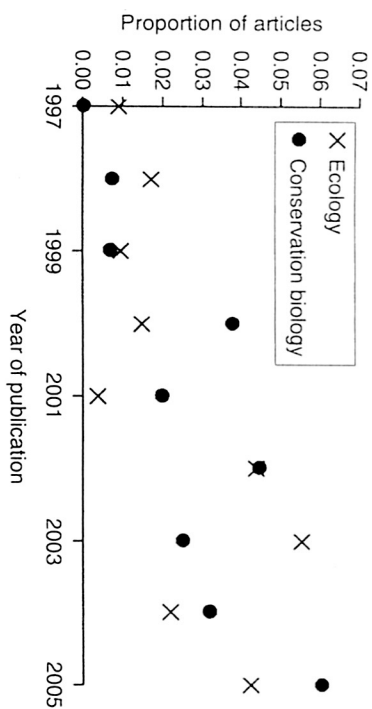


Fig. 1.1 The proportion of articles in the journals *Ecology* and *Conservation Biology* that refer to 'Bayes' or 'Bayesian'.

statistics, ignore criticisms of these methods and do not acknowledge that there are alternatives (Fowler *et al.*, 1998; Sokal and Rohlf, 1995; Underwood, 1997; Zai, 1999). Those that do address the controversy usually aim to change the status quo (Hilborn and Mangel, 1997; Burnham and Anderson, 2002), although there are exceptions (Quinn and Keough, 2002; Gotelli and Ellison, 2004).

The Bayesian approach is used relatively rarely (Fig. 1.1), so why should it interest ecologists? There are several reasons but two are particularly relevant ones. Firstly, Bayesian methods are fully consistent with mathematical logic, while conventional statistics are only logical when making probabilistic statements about data, not hypotheses (Cox, 1946; Berger and Berry, 1988; Jaynes, 2003). Bayesian methods can be used to make probabilistic predictions about the state of the world, while conventional statistics are restricted to statements about long-run averages obtained from hypothetical replicates of sampled data.

Secondly, relevant prior information can be incorporated naturally into Bayesian analyses by specifying the appropriate prior probabilities for the parameters. In contrast, conventional statistical methods are forced to ignore any relevant information other than that contained in the data. Difficulties with Bayesian methods and other benefits are discussed more fully in Chapter 2 and throughout this book.

Bayesian statistics are founded on the work of the Reverend Thomas Bayes, who lived and died in eighteenth century England (Box 1.1). Bayesian methods explicitly recognize and combine four

### Box 1.1

#### The Reverend Thomas Bayes, FRS



Very little is known about Thomas Bayes. The portrait above (O'Donnell, 1936) may be of Bayes, but no other portraits are known (Bellhouse, 2004). Even the year (1701 or 1702) and place of his birth (London or Hertfordshire, England) are uncertain (Dale, 1999). There are few records to indicate the nature of his early schooling, but he is known to have studied divinity and mathematics at the University of Edinburgh. He was ordained as a Presbyterian minister by 1728. He was elected as a Fellow of the Royal Society in 1742 but it was not until after his death in 1761 that his most famous contribution, his essay in the *Philosophical Transactions* of the Royal Society of London, was published (Bayes, 1763). In that essay, Bayes described his theory of probability and presented what is now known as Bayes' rule (or Bayes' theorem), establishing the basis of Bayesian statistics.

components of knowledge. Prior knowledge and new data are combined using a model to produce posterior knowledge.<sup>2</sup> These four components may be represented as:

$$\text{prior} + \text{data} \xrightarrow{\text{model}} \text{posterior}$$

It is common in everyday life to combine prior information and new data to update knowledge. We might hear a weather forecast that the chance of rain is small. However, if we stepped outside and saw dark

<sup>2</sup> Prior and posterior refer to before and after considering the data.

clouds looming above us, most people would think that the risk of rain was higher than previously believed. In contrast, our expectation of a fine day would be reinforced by a sunny sky. Thus, both the prior information (the weather forecast) and the data (the current state of the weather) influence our newly updated belief in the prospects of rain.

Our updated belief in the chance of rain (the posterior) will depend on the relative weight we place on the prior information compared to the new data and the magnitude of the difference between the two pieces of information. In this case the 'model' is contained within our understanding of the weather. Our thought processes combine the prior information, data, and model to update our belief that it will rain. Bayesian statistics provide a logically consistent, objective and repeatable method for combining prior information with data to produce the posterior, rather than the subjective judgement that most people would use when stepping outside.

Before considering the benefits and limitations of Bayesian methods and its alternatives in Chapter 2, I will illustrate the use of the different statistical approaches with two examples. These highlight how Bayesian methods provide answers to the kinds of questions that ecologists ask, and how they can usefully incorporate prior information.

### Example 1: Logic in determining the presence or absence of a species

Consider an ecologist who surveys ponds in a city for frogs. On her first visit to a pond, she searches the edge and listens for frog calls over a 20-minute period. The southern brown tree frog (*Litoria ewingi*) is the most common species in her study area, but it is not found on this particular visit (Fig. 1.2). However, the researcher would not be particularly surprised that the species was not detected because she knows from experience that when surveying ponds, southern brown tree frogs are detected on only 80% of visits when they are in fact present. Given this information, what can she conclude about whether the southern brown tree frog is present at the site or not?

The question about the presence of a species is a simple example of those asked by ecologists. We assume that there is a particular true state of nature and we hope to use scientific methods to determine a reasonable approximation of the truth. However, the probability that a species is



Fig. 1.2 The southern brown tree frog *Litoria ewingi*, a common species in the ponds of Melbourne, Victoria. Photograph by Nick Clemann.

present at a site is rarely calculated by ecologists, although it should be a fundamental part of any field study that depends on knowing where a species does and does not occur. This probability is not calculated partly because the statistical methods used by most ecologists are not well-suited to this question. I will examine three different approaches to answering this question and demonstrate that a satisfactory answer requires Bayesian methods.

#### Frequentist approaches

Conventional approaches to data analysis in ecology estimate the likelihood of observing the data (and more extreme data in the case of null hypothesis testing). These approaches are referred to as frequentist methods because they are based on the expected frequency that such data would be observed if the same procedure of data collection and analysis was implemented many times. Frequentist methods focus on the frequency with which the observed data are likely to be obtained from hypothetical replicates of sampling.

There are numerous types of frequentist statistics that are used in ecology, including null hypothesis significance testing and information-theoretic methods. These are applied below to the question about whether southern brown tree frogs are present at the pond.

### Null hypothesis significance testing

The first statistical approach to answering the question is null hypothesis significance testing. The null hypothesis for this first case might be that the southern brown tree frog is absent from the site. The researcher then seeks to disprove the null hypothesis with the collection of data. The single piece of data in this case is that the frog was not detected. The researcher then asks: 'What is the probability of obtaining this result if the null hypothesis were true?'<sup>3</sup> This probability is the  $p$ -value of the significance test. If the  $p$ -value is sufficiently small (conventionally if less than 0.05), it means that the data (or more extreme data) would be unlikely to occur if the null hypothesis is true. If the  $p$ -value is small, then we assume that the data are inconsistent with the null hypothesis, which is then rejected in favour of the alternative.

In the case of the frog survey, the  $p$ -value is equal to 1.0. This is calculated as the probability that we would fail to record the frog (i.e. obtain the observed data) if it is absent (i.e. if the null hypothesis is true). The high  $p$ -value means that the researcher fails to reject the null hypothesis that the frog is absent.

The other possible null hypothesis is that the frog is present at the site. In this case, the probability of obtaining the data is equal to 0.2 (one minus the probability of detecting the species if present) given that the null hypothesis is true. Thus, the  $p$ -value is 0.2, and using a conventional cut-off of 0.05, the researcher would have a non-significant result. The researcher would fail to reject the null hypothesis that the southern brown tree frog was present.

It is surprising (to some people) that the two different null hypotheses can produce different results. The conclusion about whether the species is present or absent simply depends on which null hypothesis we choose. The source of this surprise is our failure to consider statistical power, which I will return to in Chapter 2.

Another possible source of surprise is that the  $p$ -value does not necessarily provide a reliable indicator of the support for the null hypotheses. For example, the  $p$ -value is equal to 1.0 for the null hypothesis that the frog is absent. This is the largest possible  $p$ -value, but it is still not proof that the null hypothesis is true. If we continued to return to the

same pond and failed to find the frog, the  $p$ -value would remain equal to 1.0, insensitive to the accumulation of evidence that the frog is absent. This apparent discrepancy occurs because frequentist methods in general and  $p$ -values in particular do not provide direct statements about the reliability of hypotheses (Berger and Sellke, 1987; Berger and Berry, 1988). They provide direct information about the frequency of occurrence of data, which only gives indirect support for or against the hypotheses. In this way, frequentist methods are only partially consistent with mathematical logic, being confined to statements about data but not directly about hypotheses (Berger and Sellke, 1987; Jaynes, 2003).

### Information theoretic methods

An information theoretic approach based on 'likelihood' is an alternative frequentist method to null hypothesis significance testing. It evaluates the consistency of the data with multiple competing hypotheses (Burnham and Anderson, 2002). In the current example, there are only two possible hypotheses: the frog is absent ( $H_a$ ) and the frog is present ( $H_p$ ). Likelihood-based methods ask: 'What is the probability of observing the data under each of the competing hypotheses?' In this example it is the probability of not detecting the species during a visit to a site.

Unlike null hypothesis testing, likelihood-based methods, including information-theoretic methods, do not consider the possibility of more extreme (unobserved) data. The likelihood for a given hypothesis can be calculated as the probability of obtaining the data given that the hypothesis is true.<sup>4</sup> Despite the implication of its name, the likelihood of a hypothesis is not the same as the probability that the hypothesis is true. Under the first hypothesis (the frog is absent), the probability of observing the data ( $\Pr(D|H_a)$ ) is equal to 1. Under the second hypothesis (the frog is present) the probability ( $\Pr(D|H_p)$ ) is 0.2. Information-theoretic methods then determine the amount of evidence in favour of these two hypotheses by examining the ratio of these values (Burnham and Anderson, 2002).<sup>5</sup> These ratios may be interpreted by rules of thumb (see also Chapter 4). Using the criteria of Burnham and Anderson (2002),

<sup>3</sup>

In actual fact, a null hypothesis significance test asks what is the probability of obtaining the data *or a more extreme result*. However, in this case, a more extreme result is not possible; it is not possible to fail to detect the frog more than once with one visit, so the  $p$ -value is simply the probability of observing the data.

<sup>4</sup>

The likelihood need only be proportional to the probability of obtaining the data, not strictly equal to it. Terms that do not include the data or the parameters being estimated can be ignored because they will cancel out of the subsequent calculations.

<sup>5</sup>

Information-theoretic methods are modified by the number of parameters that are estimated with the data. In this case, the parameter of the analyses (the detection rate) is not estimated with the data, so the number of estimated parameters is zero.



we might conclude that the southern brown tree frog is 'considerably less' likely to be present than it is to be absent ( $\Pr(D | H_p)/\Pr(D | H_a) = 1/5$ ).

### Bayesian methods

Frequentist methods are in general not well-suited to the species detection problem because they are strictly limited to assessing long-run averages rather than predicting individual observations (Quinn and Keough, 2002). This is revealing; frequentist methods are not strictly suitable for predicting whether a species is absent from a particular site when it has not been seen. Such a problem is fundamental in ecology, which relies on knowing the distribution of species. In contrast, the species detection problem can be tackled using Bayesian methods.

Bayesian methods are similar to likelihood-based methods, but also incorporate prior information using what is known as 'prior probabilities'. Bayesian methods update estimates of the evidence in favour of the different hypotheses by combining the prior probabilities and the probabilities of obtaining the data under each of the hypotheses. The probability that a hypothesis is true increases if the data support it more than the competing hypotheses.

Why might the prior information be useful? If the researcher visited a pond that appeared to have excellent habitat for southern brown tree frogs (e.g. a large well-vegetated pond in a large leafy garden), then a failure to detect the species on a single visit would not necessarily make the researcher believe that the frog was absent. However, if the researcher visited a pond that was very unlikely to contain the frog (e.g. a concrete fountain in the middle of an asphalt car park), a single failure to detect the frog might be enough to convince the researcher that the southern brown tree frog did not occur at the pond. Frequentist methods cannot incorporate such prior information, but it is integral to Bayesian methods.

Another key difference between Bayesian methods and frequentist methods is that instead of asking: 'What is the probability of observing the data given that the various hypotheses are true?' Bayesian methods ask:

*What is the probability of the hypotheses being true given the observed data?*

At face value, this is a better approach for our problem because we are interested in the truth of the hypotheses (the frog's presence or absence at the site) rather than the probability of obtaining the observed data given different possible truths.

In practice, Bayesian methods differ from likelihood methods by weighting the likelihood values by the prior probabilities to obtain posterior probabilities. I will use the two symbols  $\Pr(H_a)$  and  $\Pr(H_p)$  to represent the prior probabilities. Therefore, the likelihood for the presence of the frog given that it was not seen (0.2) is weighted by  $\Pr(H_p)$  and the likelihood for the absence of the frog (1.0) is weighted by  $\Pr(H_a)$ . Thus, the posterior probability of presence is a function of the prior probability  $\Pr(H_p)$ , the data (the frog was not seen) and the model, which describes how the data were generated conditional on the presence or absence of the frog. Now we must determine a coherent scheme for determining the values for the prior probabilities  $\Pr(H_p)$  and  $\Pr(H_a)$ . This incorporation of prior information is one of the unique aspects of Bayesian statistics. It also generates the most controversy.

Both hypotheses might be equally likely (prior to observing the data) if half the sites in the study area were occupied by southern brown tree frogs (Parris unpublished data). In this case,  $\Pr(H_a) = 0.5$ , as does  $\Pr(H_p)$ . With these priors, the probability of the southern brown tree frog being absent will be proportional to  $0.5 \times 1.0 = 0.5$ , and the probability of it being present will be proportional to  $0.5 \times 0.2 = 0.1$ .

The posterior probabilities must sum to one, so these proportional values (0.5 and 0.1) can be converted to posterior probabilities by dividing by their sum ( $0.5 + 0.1 = 0.6$ ). Therefore, the probability of the frog being present is  $1/6 (= 0.1/0.6)$ , and the probability of absence is  $5/6 (= 0.5/0.6)$ . So, with equal prior probabilities ( $\Pr(H_a) = \Pr(H_p) = 0.5$ ), we would conclude that the presence of the frog is five times less probable than the absence of the frog because the ratio ( $\Pr(H_p | D)/\Pr(H_a | D)$ ) equals  $1/5$ . You may have noticed that this result is numerically identical to the likelihood-based result. I will return to this point later.

A different prior could have been chosen for the analysis. A statistical model predicts the probability of occupancy of ponds by southern brown tree frogs based on the level of urbanization (measured by road density), characteristics of the vegetation, and the size of the pond (based on Parris 2006.). If the pond represented relatively high-quality habitat, with a predicted probability of occupancy of 0.75, then the probability of the frog being present will be proportional to  $0.75 \times 0.2 = 0.15$  and the probability of absence will be proportional to  $(1 - 0.75) \times 1.0 = 0.25$ . With these priors, the probability of the frog being present is equal to  $3/8 (= 0.15/(0.15 + 0.25))$ , and the probability of absence is  $5/8 (= 0.25/(0.15 + 0.25))$ .

The incorporation of prior information (the presence of good quality habitat) increases the probability that the pond is occupied by southern brown tree frogs compared to when the prior information is ignored (0.375 versus 0.167). The actual occupancy has not changed at all – the pond is still either occupied or not. What has changed is the researcher's belief in whether the pond is occupied. These Bayesian analyses may be formalized using Bayes' rule, which, following a short introduction to conditional probability (Box 1.2), is given in Box 1.3.

### Box 1.2

#### Conditional probability

Bayes' rule is based on conditional probability. Consider two events: event  $C$  and event  $D$ . We are interested in the probability of event  $C$  occurring given event  $D$  has occurred. I will write this probability using the symbol  $\Pr(C|D)$ , and introduce three more symbols:

$\Pr(C)$  – the probability of event  $C$  occurring;

$\Pr(D)$  – the probability of event  $D$  occurring; and

$\Pr(C \text{ and } D)$  – the probability of both events occurring together.

Conditional probability theory tells us that:

$$\Pr(C \text{ and } D) = \Pr(D) \times \Pr(C|D),$$

which in words is: the probability of events  $C$  and  $D$  both occurring is equal to the probability of event  $C$  occurring given that event  $D$  has occurred multiplied by the probability of event  $D$  occurring (independent of event  $C$ ). The  $|$  symbol means 'given the truth or occurrence of'.

The above can be rearranged to give:

$$\Pr(C|D) = \Pr(C \text{ and } D) / \Pr(D).$$

For example, *Pfiesteria*, a toxic alga is present in samples with probability 0.03 (Slow and Borsuk 2003). *Pfiesteria* is a subset of *Pfiesteria*-like organisms (PLOs), the latter being present in samples with probability 0.35. Therefore, we can calculate the conditional probability that *Pfiesteria* is present given that PLOs are present:

$$\begin{aligned}\Pr(Pfiesteria|PLO) &= \Pr(Pfiesteria \text{ and } PLO) / \Pr(PLO) \\ &= 0.03 / 0.35 = 0.086.\end{aligned}$$

### Box 1.3

#### Bayes' rule for a finite number of hypotheses

Conditional probability (Box 1.2) states that for two events  $C$  and  $D$ :

$$\Pr(C \text{ and } D) = \Pr(D) \times \Pr(C|D).$$

$C$  and  $D$  are simply labels for events (outcomes) that can be swapped arbitrarily, so the following is also true:

$$\Pr(D \text{ and } C) = \Pr(C) \times \Pr(D|C).$$

These two equivalent expressions for  $\Pr(C \text{ and } D)$  can be set equal to each other:

$$\Pr(D) \times \Pr(C|D) = \Pr(C) \times \Pr(D|C).$$

It is then straightforward to obtain:

$$\Pr(C|D) = \Pr(C) \times \Pr(D|C) / \Pr(D).$$

Let us assume that event  $C$  is that a particular hypothesis is true, and event  $D$  is the occurrence of the data. Then, the posterior probability that the frog is absent given the data ( $\Pr(H_a|D)$ ) is:

$$\Pr(H_a|D) = \Pr(H_a) \times \Pr(D|H_a) / \Pr(D).$$

The various components of the equation are the prior probability that the frog is absent ( $\Pr(H_a)$ ), the probability of obtaining the data given that it is absent ( $\Pr(D|H_a)$ , which is the likelihood), and the probability of obtaining the data independent of the hypothesis being considered ( $\Pr(D)$ ).

The probability of obtaining the data (the frog was not detected) given  $H_a$  is true (the frog is absent) was provided when using the likelihood-based methods:

$$\Pr(D|H_a) = 1.0.$$

Similarly, given the presence of the frog:

$$\Pr(D|H_p) = 0.2.$$

The value of  $\Pr(D)$  is the same regardless of the hypothesis being considered ( $H_p$ , the frog is present, or  $H_a$ , the frog is absent), so it simply acts as a scaling constant. Therefore,  $\Pr(H_a|D)$  is proportional to  $\Pr(H_a) \times \Pr(D|H_a)$ , and  $\Pr(H_p|D)$  is proportional to

$\Pr(H_p) \times \Pr(D | H_p)$ , with both expressions having the same constant of proportionality ( $1/\Pr(D)$ ).

$\Pr(D)$  is calculated as the sum of the values  $\Pr(H) \times \Pr(D | H)$  under all hypotheses. When prior probabilities are equal

( $\Pr(H_a) = \Pr(H_p) = 0.5$ ):

$$\begin{aligned}\Pr(D) &= [\Pr(H_a) \times \Pr(D | H_a)] + [\Pr(H_p) \times \Pr(D | H_p)] \\ &= (0.5 \times 1) + (0.5 \times 0.2) = 0.6.\end{aligned}$$

Therefore, the posterior probabilities are  $5/6$  ( $0.5/0.6$ ) for the absence of the frog, and  $1/6$  ( $0.1/0.6$ ) for the presence of the frog.

So, for a finite number of hypotheses, Bayes' rule states that the probability of the hypothesis given the data is calculated using the prior probabilities of the different hypotheses ( $\Pr(H_j)$ ) and the probability of obtaining the data given the hypotheses ( $\Pr(D | H_j)$ ):

$$\Pr(H_i | D) = \frac{\Pr(H_i) \times \Pr(D | H_i)}{\sum_j \Pr(H_j) \times \Pr(D | H_j)}$$

This expression uses the mathematical notation for summation  $\sum$ .

If on the other hand, the pond had poor habitat for southern brown tree frogs, the prior probability of presence might be 0.1. Thus,  $\Pr(H_p) = 0.1$  and  $\Pr(H_a) = 0.9$ . As before,  $\Pr(D | H_p) = 0.2$  and  $\Pr(D | H_a) = 1.0$ . Note that the values for the priors but not the likelihoods have changed. Using Bayes' rule (Box 1.3), the posterior probability of presence is:

$$\begin{aligned}\Pr(H_p | D) &= \Pr(H_p) \times \Pr(D | H_p) / [\Pr(H_p) \times \Pr(D | H_p) + \Pr(H_a) \times \Pr(D | H_a)] \\ &= 0.1 \times 0.2 / [0.1 \times 0.2 + 0.9 \times 1.0] \\ &= 0.022\end{aligned}$$

Therefore, there is only a small chance that the frog is at the site if it has poor habitat and the species is not detected on a single visit.

Bayesian methods use probability distributions to describe uncertainty in the parameters being estimated (see Appendix B for more background on probability distributions). Probability distributions are used for both priors and posteriors. The frog surveying problem has

two possible outcomes; the frog is either present or absent. Such a binary outcome (e.g., presence/absence, heads/tails, increasing/decreasing) can be represented by a Bernoulli probability distribution, which is a special case of the binomial distribution with a sample size of one. Bernoulli random variables take a value of one (representing the presence of the frog) with a probability equal to  $p$  and a value of zero (representing the absence of the frog) with probability  $1-p$ . Therefore, uncertainty about the presence of the frog at the pond can be represented as a Bernoulli random variable in which the probability of presence is equal to  $p$ .

It is important to note that a probability distribution is used to represent the *uncertainty* about the presence of the frog. The frog is assumed to be actually present or absent at the site, and the distribution is used to represent the probability that it is present. There appears to be misunderstanding among at least some ecologists that Bayesian parameters do not have fixed values, but change randomly from one measurement to another. Although such models can be accommodated within Bayesian analyses (e.g., by using hierarchical models, Box 3.6), parameters are usually assumed to have fixed values. The prior and posterior distributions are used to represent the uncertainty about the estimate of the parameters.

I have illustrated three components of a Bayesian analysis: priors, data and posteriors. I have not explicitly stated the model, which is the fourth aspect I mentioned in the introduction. The model in the above example is relatively simple and is the same as was used in the frequentist analyses. It can be stated as: 'the detection of the southern brown tree frog during the survey occurs randomly with a probability ( $p_{\text{detect}}$ ) that depends on whether the pond is occupied ( $p_{\text{detect}} = 0.8$ ) or not ( $p_{\text{detect}} = 0.0$ )'.

This model may be written algebraically as:

$$\begin{aligned}p_{\text{detect}} &= 0.8 \times \text{present} \\ \text{detected} &\sim \text{Bernoulli}(p_{\text{detect}}).\end{aligned}$$

The second expression says that the variable called 'detected' is a Bernoulli random variable. A value of one for 'detected' indicates that the frog was detected and a zero indicates it was not. The probability of detection is equal to  $p_{\text{detect}}$ , and is given in the first equation. It depends on whether the frog is present at the site ( $\text{present} = 1$ ,  $p_{\text{detect}} = 0.8$ ) or absent ( $\text{present} = 0$ ,  $p_{\text{detect}} = 0.0$ ).

## Random sampling from the posterior distribution using WinBUGS

This Bayesian analysis can also be implemented in the freely available software package WinBUGS (Spiegelhalter *et al.*, 2005). Appendix A provides information about obtaining the program WinBUGS and a tutorial on its use. I will use WinBUGS throughout the book, so it is worth investing some time in understanding it. Readers who are unfamiliar with WinBUGS should study Appendix A now, before continuing with the rest of the book.

The acronym WinBUGS is based on the original program BUGS (Bayesian inference Using Gibbs Sampling), but is now designed to run under the Microsoft Windows operating system (hence the Win prefix). WinBUGS works by randomly sampling the parameters used in Bayesian models from their appropriate posterior distributions. Because the posterior distribution for the example of detecting southern brown treefrogs can be calculated (Box 1.3), it is not necessary to use WinBUGS in this case. However, for many problems it is difficult or impossible to calculate the posterior distribution, but samples from it can be obtained relatively easily using WinBUGS or other MCMC software. If a sufficiently large number of replicates are taken, the form of the posterior distribution can be determined and its parameters, such as the mean, standard deviation, and percentiles, can be estimated.

WinBUGS takes samples from the posterior distribution by using 'Markov chain Monte Carlo' (MCMC) methods. 'Monte Carlo' implies random sampling, referring to roulette wheels and other games of chance. 'Markov chain' refers to the method of generating the random samples. A series of random numbers in which the value of each is conditional on the previous number is known as a Markov chain. MCMC algorithms are constructed in such a way that the samples from the Markov chain are equivalent to samples from the required posterior distribution (see Appendix C).

The advantage of using Markov chains for sampling from the posterior distribution is that it is not necessary to calculate the value of the denominator in Bayes' rule. The calculation is avoided because each successive sample depends on the ratio of two posterior probabilities that share the same denominator, which then cancels (Appendix C). This simplifies matters, because the Bayesian analysis only requires the product of the prior probability and the likelihood of the data.

If each sample depends on the value of the previous sample, successive values drawn from the Markov chain may be correlated. Correlations between the samples have some important consequences. The first is that the initial values that are used in the Markov chain may influence the results until a sufficiently large number of samples is generated. After this time, the 'memory' of the initial values is sufficiently small and the samples will be drawn from the posterior distribution (Box 1.4). Because of the potential for dependence on the initial values, their possible influence is

### Box 1.4

#### The burn-in when sampling from Markov chains

It can take thousands of iterations for some Markov chains to converge to the posterior distribution, while others converge immediately. Therefore, it is necessary to check convergence, and discard the initial samples from a Markov chain until convergence is achieved. These discarded values are referred to as a 'burn-in'.

There are several ways to check for convergence. One of the simplest is to plot the sampled values versus the iteration number. In the example in Fig 1.3, the initial value is approximately 1200, changing to values in the approximate range 100 to 400 after five samples. The values continue to be around 100 to 400 indefinitely, suggesting that the chain has reached what is known as its stationary distribution. The Markov chain is constructed in such a way for Bayesian analyses that this stationary distribution is the posterior distribution (Appendix C).

A further check for stationarity is to initiate the Markov chain with a second set of initial values. The stationary distribution will be insensitive to the initial values. If two chains with different initial values converge, then it suggests that both chains have reached their stationary distribution. There are formal methods for checking the convergence of a pair of Markov chains, such as the Gelman-Rubin statistic (Brooks and Gelman, 1998), which compares the variation of the samples within chains and the variation of the samples when the chains are pooled. Initially, the pooled variation will be greater than the average variation of the two chains and then become equal as the chains converge. Additionally, the level of variation both within and between chains should stabilize with convergence.

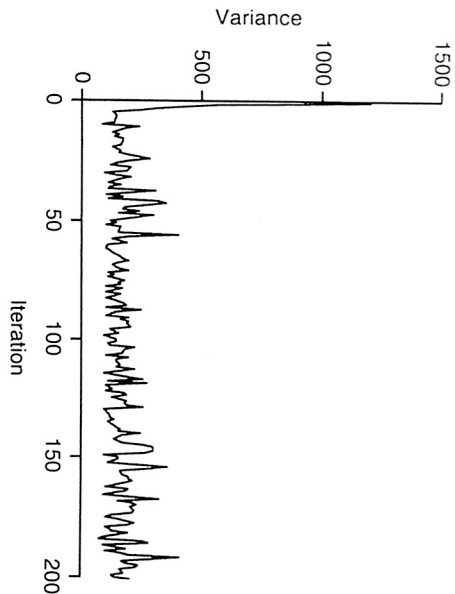


Fig. 1.3 The first 200 samples of the variance of the number of trees in a remnant for the model in Box 3.2.

examined and it may be necessary to discard some of the initial samples (perhaps the first few thousand or more) as a 'burn in' (Box 1.4).

A second consequence of any correlation is that, compared to an uncorrelated sample, each additional sample contains only a fraction of the information about the posterior distribution. Because of this, a large number of samples may be required to obtain a sufficiently precise sample if there is strong correlation between samples. Although the presence of correlation in the Markov chain reduces the efficiency of the sampling algorithm, it does not preclude the use of Markov chain methods. The reduced efficiency is simply the cost to be paid when it is not possible to obtain an analytical solution for the posterior distribution. Gilks *et al.* (1996) provides further information about Markov chain Monte Carlo methods.

### The frog surveying problem in WinBUGS

Code for analysing the frog surveying problem in WinBUGS is given in Box 1.5. A Bayesian model specified in WinBUGS has the four components of a Bayesian analysis:

- prior distributions for the parameters being estimated;
- data;
- a model that relates the parameters to the data; and
- the posterior distributions for the parameters.

### Box 1.5 WinBUGS code for determining the presence of a species

The frog surveying problem involves determining whether the species is present at a site given that it was not detected during a survey. In WinBUGS, the code works by specifying the prior for the probability of presence and the model, which describes how the parameter of interest (the presence of the frog) is related to the data. Pseudo-code for this problem would be:

1. Specify the prior probability of presence;
2. Specify that the frog is either present or absent with a particular probability;
3. Calculate the probability of detecting the species given that it is either present (probability of detection = 0.8) or absent (probability of detection = 0.0);
4. Specify that the detection of the frog or failure to detect the frog (the data) arises randomly, depending on the probability of detection.

Steps 1–2 specify the prior for the presence of the frog. Steps 3–4 specify the model, describing how the data (the observation of an absence in this case) are related to the presence of the frog, which is the parameter being estimated.

The WinBUGS code for the frog surveying problem is written below.

```
model
{
  prior <- 0.5

  # the prior
  # probability of
  # presence
  present ~ dbern(prior)
  # actual presence
  # drawn from a
  # Bernoulli
  # dist'n
  prob_detect <- 0.8*present
  # prob of
  # detection depends
  # on
  # presence/absence
```



```

detected ~ dbern(prob_detect) # actual detection
occurs with
random variation
}
list(detected = 0)

# the data - the frog
was not detected

```

In this model we are interested in determining whether the frog is present (represented by the variable `present`). The variable prior is the prior probability of the frog being present. The prior probability of the frog being absent is therefore `1 - prior`. The actual presence at the site is determined randomly, by drawing from a Bernoulli distribution; a value of one indicates the frog is present and zero indicates the frog is absent. Therefore, the first two lines define the expected presence of the frog prior to the collection of the data.

The next two lines describe the model of how the data were collected. If present, the probability of detecting the frog (`prob_detect`) is equal to 0.8, and it will equal zero if it is absent. The fourth line then states that the data are assumed to occur randomly, again drawn from a Bernoulli distribution, with the probability of detecting the frog on a single visit being equal to `prob_detect`, and the probability of not detecting the frog being equal to `1 - prob_detect`.

The observed data (written in the line `list(detected = 0)`) then influence the values of the variable `present`, through the application of Bayes' rule within WinBUGS (Box 1.3). Values of the variable `present` are sampled by WinBUGS such that they are drawn as random samples from its posterior distribution. Sampling in this way is called Monte Carlo sampling. It is a relatively common method of analysing probabilistic models (Box 1.6). If enough samples are taken, the probability of the frog being present can be estimated by the proportion of times that the variable `present` equals one. This proportion equals the mean of the variable `present`.

Sampling 100 000 times from this model in WinBUGS (after ignoring the first 10 000 samples) leads to a mean value of `present` of 0.17, which is equivalent to 1/6, as determined analytically. This is our estimate of the posterior probability that the site is occupied given the prior and the data. Changing the value of `prior` to 0.75 leads to a mean value of `present` that is equal to 0.38 (again based on 100 000 samples), which is equivalent to 3/8 as determined analytically.

The results in WinBUGS are not exact because of random sampling error. If we took more samples in WinBUGS, the results would be closer to the truth. For example, the posterior probability of presence equals 0.3754 if half a million samples are taken when the prior for this value is 0.75. It is not precisely the same as the true answer (0.3750), but the answer in WinBUGS will continue to become more precise as more samples are taken.

### Box 1.6

#### Monte Carlo methods

Monte Carlo methods use simulation to estimate the probability of occurrence of uncertain events. For example, consider a five-card poker hand. We could use probability theory to work out the chance of obtaining a flush (five cards of the same suit). The probability is equal to the probability that the second, third, fourth and fifth cards are the same suit as the first. For a 52-card deck, this is equal to:

$$(12/51) \times (11/50) \times (10/49) \times (9/48) = 0.00033$$

We could also work out this probability with a Monte Carlo method by dealing, shuffling, and re-dealing and calculating the proportion of times that a flush appears. If we did this ten times, we might get one flush (if we were lucky). Based on these results (one occurrence out of ten deals), we might estimate that the probability of a flush is 0.1. This is an imprecise estimate. Obtaining more samples increases the precision. If we dealt the cards 10 000 times, we might get three flushes, implying that the probability of a flush is 0.0003. This is better, but still not perfect; we could deal the cards several million times and get an even more precise estimate of the probability.

Of course, it is laborious to deal the cards that many times. An efficient alternative might be for a machine to deal the cards for us.

Such a task might be suitable for computers, because they specialize in repetitive tasks. However, instead of dealing a physical deck of cards, the computer could use its circuitry to generate 'random' numbers that have the same statistical properties as the cards. In this case, thousands of samples can be generated very quickly by randomly generating integers between 1 and 52 (representing the 52 possible cards) with equal probability.

This virtual random sampling is the same sort of process that is used by WinBUGS. It generates samples that have the same statistical properties as the posterior distribution. The samples generated by WinBUGS can then be analysed to estimate the statistical properties of the posterior distribution such as its mean and percentiles.

WinBUGS code includes the prior for the parameters, but most of the code is usually the model, which describes how the data are related to the parameters. The posterior is then generated by WinBUGS with Monte Carlo sampling (Box 1.6; Appendix C).

The advantage of using a Monte Carlo approach is that it is able to sample from the posterior distribution without analysts having to do the various calculations themselves. In the frog surveying problem, the calculations done by hand are relatively easy. In the few cases where the calculations can be done by hand, they are usually more difficult, and in most other cases they are impossible.

Monte Carlo methods have another appealing property. Even relatively complex statistical analyses (e.g. regression analysis) do not require WinBUGS code that is much more complex than that presented in Box 1.5. Once familiar with relatively simple analyses, it is not much more difficult to write code for more complex analyses.

## Example 2: Estimation of a mean

The second example of Bayesian analysis involves estimating the average diameter of trees in a remnant patch of eucalypt forest (Harper *et al.*, 2005). The size of trees is important when studying, for example, nutrient dynamics, provision of habitat for animals, production of nectar, mitigation of temperature extremes, and amelioration of pollution (Bormann and Likens, 1979; Attiwill and Leeper, 1987; Huang, 1987; McPherson *et al.*, 1998; Brack, 2002; Gibbons and Lindenmayer, 2002; Breton *et al.*, 2004).

The mean diameter of trees could conceivably take any value between zero and some large number. Therefore, the hypotheses are not discrete. There are an infinite number of hypotheses, represented by any conceivable value for the mean diameter of the trees. Bayesian methods are able to accommodate these sorts of cases where hypotheses are distributed

along a continuum, by using continuous rather than discrete probability distributions to represent uncertainty in the variables. The only modification to Bayes' rule is how the constant of proportionality is calculated (Box 1.7).

Assume that a researcher has measured the diameter of 10 randomly selected trees. In analysing the data, the researcher must choose the prior probability distribution for the parameters being estimated. Although the mean size of trees could be conceivably any positive number, the researcher has previously measured more than 2500 trees in 43 other

### Box 1.7 Bayes' rule for continuous hypotheses

In the case of continuous hypotheses, continuous probability distributions are used to represent different possible values for parameters. Bayes' rule is then expressed as:

$$\Pr(H | D) = \frac{\Pr(H) \times \Pr(D | H)}{\int_0^{\infty} \Pr(x) \times \Pr(D | x) dx},$$

where  $H$  represents a particular value for the parameter. The integral in the denominator substitutes for the summation in the discrete case, and the limits of the integration are over all the possible values of the parameter ( $x$ ), which in this case is assumed to be positive. This integral makes Bayesian methods difficult to conduct analytically, because in most cases it cannot be determined.

Readers who are uncomfortable with mathematics may look at the above equation and decide that they can never solve those sorts of problems and decide that Bayesian methods are too hard. The complexity of the equation should not be discouraging because in most cases it is impossible to solve, regardless of a person's mathematical skills. Fortunately, software is available so users do not need to evaluate or even construct the integral.

As with the case when there were a finite number of hypotheses (Box 1.3), the denominator simply acts as a scaling constant, because it is the same for all possible values for the parameter  $H$ . As with discrete hypotheses, the posterior probability is simply proportional to the prior probability ( $\Pr(H)$ ) multiplied by the likelihood ( $\Pr(D | H)$ ). The main analytical task of Bayesian analyses is to determine the constant of proportionality.

remnants (Harper *et al.*, 2005). After measuring so many trees in the study area, he has a good idea about likely values for the mean diameter of trees in the previously unmeasured remnant. Frequentist analyses do not permit this additional information to be used in determining the mean diameter of trees in the new remnant, but a Bayesian analysis does.

Based on data from the other 43 remnants, the mean diameter of trees in remnants is 53 cm and the mean varies among remnants with a standard deviation of approximately 5 cm. Assuming the mean diameter of trees follows a normal distribution, we would expect approximately 95% of remnants to have a mean tree diameter that is within 1.96 standard deviations of the overall average. Therefore, prior to collecting the data there is a 95% chance that the mean diameter of trees in the new remnant will be between approximately 43 and 63 cm. This prior reflects the researchers' expectation of the mean size of trees in a newly measured remnant based on his previous experience in the study area. A plot of the prior shows the range of likely values (Fig. 1.4).

### The Bayesian solution for the normal mean

In the simplest case, and to make the analysis comparable to a traditional frequentist analysis, we will assume that the diameter of trees within the

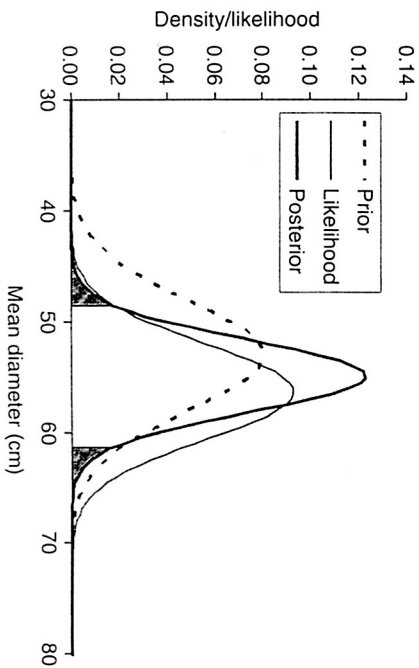


Fig. 1.4 The prior and posterior density functions and likelihood for the mean diameter of trees in a remnant, based on a sample of ten trees. The posterior would equal the likelihood if the prior was uninformative. The posterior is more precise than both the prior and the likelihood function because the posterior combines the information in both. The limits of the 95% credible interval of the posterior have 2.5% of the area under the curve in each tail (shaded).

remnant follows a normal distribution. In the case where the data and the prior both have normal distributions, Bayes' rule (Box 1.7) provides an analytical solution for the posterior distribution. However, analytical solutions are available for only a handful of Bayesian models, so I will first illustrate this example using WinBUGS (Box 1.8). It is simply a matter of specifying a prior distribution for the mean of the diameter

### Box 1.8 Estimating a mean for a normal model using WinBUGS

In estimating the mean diameter of trees, the prior has a mean of 53 cm and a standard deviation of 5 cm. In WinBUGS, the width of a normal distribution is expressed using the precision ( $1/\text{variance} = 1/\text{sd}^2$ ), which in this case is equal to  $0.04$  ( $1/25 = 1/5^2$ ).

In this example, the variance of the data is assumed to be known, making it equivalent to using a z-value rather than a t-value in a frequentist analysis. However, uncertainty in estimating the precision of the data can be included easily in the WinBUGS analysis (Chapter 3).

The pseudo-code for the WinBUGS analysis is:

1. Specify the prior for the mean diameter of trees in the remnant as being normally distributed with a mean of 53 and precision of 0.04 (standard deviation of 5).
2. Calculate the standard deviation of the data.
3. Specify the precision of the data as the inverse of the variance of the diameter of trees in the remnant (the variance equals 184.9 in this example).
4. For each of the ten trees that were measured, assume that their diameter is drawn from a normal distribution with the mean and precision as specified in steps 1 and 3.

The WinBUGS code is:

```
model
{
  m ~ dnorm(53, 0.04)      # prior for mean
  stdev <- sd(Y[])         # calculate std deviation
                             of data
}
```

```

prec <- 1/(stdev*stdev) # precision of the data =
                        1/variance
for (i in 1:10)        # for each of the ten
                        trees ...
{
  Y[i] ~ dnorm(m, prec) # diameter drawn from a
                        normal distribution
}
list(Y = c(42, 43, 58, 70, 47, 51, 85, 63, 58, 46))

```

The 'for loop', designated by the line `for (i in 1:10)` and subsequent line within the curly brackets, is equivalent to ten lines of code, one for each of the ten trees, i.e. `Y[1] ~ dnorm(m, prec)`, up to `Y[10] ~ dnorm(m, prec)`. It is shorthand to replace repetitive sections of code.

The data are provided in the line:

```
list(Y = c(42, 43, 58, 70, 47, 51, 85, 63, 58, 46))
```

The 'c' before the brackets indicates that the following data are concatenated (linked together) into the one variable, with the first variable represented by `Y[1]`, the second by `Y[2]`, etc.

This analysis also requires that the user specifies an initial value of ~~mean~~ for the Markov chain (Box 1.4). The choice is not important because the chain converges quickly to the posterior distribution in this case, and could be generated randomly.

However, in some cases the speed of convergence is increased if the Markov chain is initiated with values that are close to the posterior distribution, so the following arbitrary value was used:

```
list(m = 55) # an arbitrary initial value
```

After discarding the first 1000 samples as a burn-in, 100 000 samples were generated in WinBUGS. Of these samples, 2.5% were less than 48.5 and 2.5% were more than 61.3. Therefore the 95% credible interval is 48.5–61.3 cm for the mean diameter of trees in the remnant. This result is insensitive to the choice of the initial value.

of trees in the remnant, and then constructing a model in which the measured diameters are drawn from a distribution with that mean. The posterior distribution calculated in WinBUGS is the same as that obtained using the analytical solution (Box 1.9).

### Confidence intervals and credible intervals

A frequentist analysis would ignore the prior information and simply use the mean of the data and the standard error ( $=\sqrt{(184.9/10)}=4.3$ ),

#### Box 1.9

#### Estimating a mean for a normal model analytically

When the data and prior have normal distributions, the posterior distribution also has a normal distribution, the mean and variance of which depends, not surprisingly, on the mean and variance of the prior. The posterior distribution also depends on the sample size, mean and variance of the data. The mean and variance of the posterior can be calculated from the following formulae (Gelman *et al.* 2004):

$$\mu_{\text{post}} = \frac{\mu_{\text{prior}}/\sigma_{\text{prior}}^2 + \mu_{\text{data}}n/\sigma_{\text{data}}^2}{1/\sigma_{\text{prior}}^2 + n/\sigma_{\text{data}}^2}, \text{ and}$$

$$\sigma_{\text{post}}^2 = \frac{\sigma_{\text{prior}}^2\sigma_{\text{data}}^2/n}{\sigma_{\text{data}}^2/n + \sigma_{\text{prior}}^2},$$

where  $n$  is equal to the sample size,  $\sigma_{\text{prior}}^2$ ,  $\sigma_{\text{data}}^2$ , and  $\sigma_{\text{post}}^2$  are the variances of the prior, data and posterior, and  $\mu_{\text{prior}}$ ,  $\mu_{\text{data}}$  and  $\mu_{\text{post}}$  are the means of the distributions.

These formulae provide useful insights into Bayesian statistics.

The mean of the posterior is a weighted average of the means of the prior and data. The weights are the precisions of the prior ( $1/\sigma_{\text{prior}}^2$ ) and the data ( $n/\sigma_{\text{data}}^2$ ). The influence of the data and prior on the posterior mean depends on which is more informative. When there are no data ( $n=0$ ), the mean of the posterior is equal to the mean of the prior. When the variance of the prior is very large,  $1/\sigma_{\text{prior}}^2$  approaches zero and the mean of the posterior will be close to the mean of the data. The prior is said to be uninformative when the



posterior is influenced exclusively by the data. This is achieved by using a prior with a large variance.

The variance of the posterior has similar properties, but these are most obvious when its formula is re-arranged to be expressed as the inverse of the variance:

$$\frac{1}{\sigma_{\text{post}}^2} = \frac{n}{\sigma_{\text{data}}^2} + \frac{1}{\sigma_{\text{prior}}^2}$$

The inverse of the variance measures precision. Large values for the precision mean the variance is small. The quantity  $n/\sigma_{\text{data}}^2$  is the inverse of the standard error squared, and it measures precision in an ordinary frequentist analysis. Therefore, the precision of the posterior is simply equal to the precision based on the data (the inverse of the standard error squared) plus the precision of the prior. The precision of an estimate is increased by using prior information.

The diameter measurements of ten trees in the new remnant (42, 43, 58, 70, 47, 51, 85, 63, 58, 46 cm) have a mean of 56.3 and variance of 184.9. Given the prior has a mean and variance of 53 and 25, the posterior distribution for the mean diameter of trees in the new remnant has the following mean and variance:

$$\mu_{\text{post}} = \frac{53/25 + 56.3 \times 10/184.9}{1/25 + 10/184.9} = 54.9$$

$$\sigma_{\text{post}}^2 = \frac{25 \times 184.9/10}{184.9/10 + 25} = 10.6$$

The standard deviation of the posterior is 3.26 cm ( $\sqrt{10.6}$ ). Therefore, there is an approximate 95% chance that the mean diameter of trees in the park is between 48.5 cm and 61.3 cm (the mean of the posterior plus or minus 1.96 times the standard deviation of the posterior) (Fig. 1.4). This 95% credible interval is the same as that obtained from WinBUGS (Box 1.8).

leading to a 95% confidence interval of 47.9–64.7 cm ( $56.3 \pm 1.96 \times 4.3$ ). This is the same as the credible interval that was obtained when using a Bayesian analysis with an uninformative prior (Box 1.10).

Bayesian credible intervals and frequentist confidence are usually numerically identical if the Bayesian prior is uninformative. An

### Box 1.10 Estimating the mean of a normal model with an uninformative prior

An uninformative prior for the mean diameter of trees can be specified by using the following line of code for the prior instead of the one in Box 1.8:

mean ~ dnorm(0, 1.0E-6) # wide prior for mean

This is a very wide normal distribution with a mean of zero and a standard deviation of 1000. Therefore, mean diameters between, for example, zero and 200 cm have approximately the same prior probability. If this uninformative prior is used, the posterior distribution for the mean diameter of trees in the remnant has a mean of 56.3 and 95% credible interval of 47.8–64.7, numerically equivalent to the 95% confidence interval of a frequentist analysis.

uninformative prior is one in which the data (via the likelihood, which is  $\Pr(D|H)$  in Bayes' rule) dominates the posterior. This is achieved by using a prior with a large variance. A large variance permits the parameter to be drawn from a wide range of possible values and the prior probabilities of all reasonable parameter values are approximately equal. When the prior distribution is uninformative, the posterior distribution has the same form as the likelihood (Fig. 1.5). The likelihood and posterior have different forms when the prior is informative (Fig. 1.4). The posterior distribution is less precise, and hence the credible interval is wider, if the prior information is ignored (Fig. 1.5). Ignoring the prior information would imply that the researcher believed that the remnant could have any mean diameter prior to collecting the data. Such a belief would be inconsistent with the researchers' previous experience in the study area, which provides useful data on the range of likely results.

Although frequentist confidence intervals and Bayesian credible intervals may appear similar, they are in fact different. For a 95% Bayesian credible interval, there is a 95% chance that the true value of the parameter will be within the interval. Ecologists are often interested in this kind of interval because they want to know the chance that the true value of the parameter is within a specified range. Such an answer requires the use of Bayesian credible intervals.



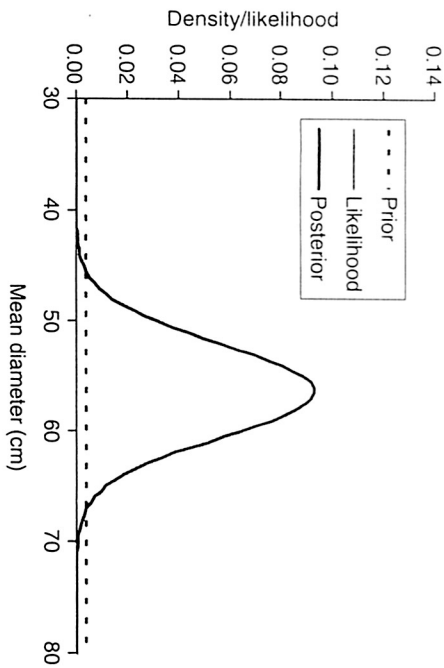


Fig. 1.5 The prior and posterior density functions and likelihood for the mean diameter of trees in a remnant, based on a sample of ten trees and using an uninformative prior. The prior distribution (drawn on an arbitrary scale to assist comparison) has a mean of zero and standard deviation of 1000 making mean diameters between 30 and 80 cm all equally likely a priori. The likelihood and the posterior are indistinguishable. The posterior is less precise than in Fig. 1.4 because the prior is uninformative.

In contrast, a 95% frequentist confidence interval does not contain the true parameter with 95% probability. Instead, it is based on the concept of an infinite number of samples. If I repeat the data collection an infinitely large number of times and construct 95% confidence intervals for the mean for each set of data, 95% of these confidence intervals would encompass the true mean.

This different meaning of confidence and credible intervals is not just semantic. In some circumstances, it can lead to numerical differences even when the credible interval is based on an uninformative prior. For example, in estimating a fail-safe period from three observations of failure times (12, 14 and 16), Jaynes (1976) shows that the shortest possible 90% confidence interval is 12.1–13.8. This interval does not contain the true fail-safe period, which must be less than the smallest observed lifespan (12). This result is not an error. The method of calculating 90% confidence intervals will produce intervals that enclose the true value of the parameter 90% of time. However, the true value might surely lie outside any single interval, as in this example.

In contrast, the Bayesian analysis with an uninformative prior arrives at a sensible conclusion; the shortest possible 90% credible interval is 11.2–12.0 (Jaynes, 1976). When the intervals are the same, the choice of

Bayesian or frequentist methods does not matter. However, when the intervals are different, only Bayesian methods provide logical results (Jaynes, 1976).

### Concluding remarks

In introducing Bayesian methods, this chapter made two important points. Firstly, Bayesian methods can answer questions that are relevant to ecologists, such as: 'What is the probability that this hypothesis is true?' and 'What is the probability that a parameter will take values within a specified interval?' Secondly, relevant prior information can also be incorporated into Bayesian analyses to improve the precision of estimates.

Bayes' rule is the basis of Bayesian methods. It is derived as a simple expression of conditional probability. The rule specifies how prior information and data are combined using a model to arrive at the posterior state of knowledge. Both the prior and posterior states of knowledge are represented as probability distributions. The posterior probability simply equals the prior probability multiplied by the likelihood of the data and a scaling constant. Bayesian methods become difficult because the scaling constant is usually hard to calculate analytically. However, recent numerical methods such as Markov chain Monte Carlo make Bayesian methods accessible to all scientists.

Frequentist confidence intervals and Bayesian credible intervals will usually be numerically equivalent if uninformative priors are used. In this way Bayesian methods provide a numerical generalization of frequentist methods. They also do so in such a way that probabilistic statements about the state of nature are mathematically logical. The next chapter provides a more thorough comparison of different statistical schools and examines their various strengths and weaknesses.